Misic T., Gilani S. and McArthur J. (2020). "BAS data streaming for Smart Building analytics" In: *Proc. 37th CIB W78 Information Technology for Construction Conference* (CIB W78), São Paulo, Brazil, pp. 430-439. DOI: http://dx.doi.org/10.46421/2706-6568.37.2020.paper031

BAS DATA STREAMING FOR SMART BUILDING ANALYTICS

Tony Misic[1], Sara Gilani[2] and J.J. McArthur[3]

Abstract: Many existing and new buildings are equipped with building automation system (BAS). BAS-integrated sensors continuously monitor environmental conditions, energy use, HVAC and lighting systems, and occupancy in buildings, collecting vast amounts of data that can be of great value for building performance optimization from both the energy use and occupant comfort perspectives. However, the heterogeneity and volume of these data pose significant barriers to their use. For the effective analysis of the collected BAS data to facilitate actionable use of them and support smart buildings, advanced analytics methods such as artificial intelligence should be deployed in real time or near real time, requiring a coherent data management strategy (streaming, pre-processing, and structuring) and integration with advanced analytics techniques. A case study whereby BAS data are collected in an academic building in Toronto, Canada, is streamed to a cloud-hosted research platform Using the BACnet software, data acquired by various sensors are collected by a BAS and streamed as tuples through a Virtual Private Network (VPN) to the cloud using Transmission Control Protocol/Internet Protocol (TCP/IP) packet messages to ensure information security. The destination of the information was an ElasticSearch (ES) cluster, which is also used as a search and analytics engine on the back end. The data streaming, pre-processing, and structuring into an ontology to support facility management and complex event processing is described in this paper along with insight regarding the stakeholder planned uses and expected benefits.

Keywords: Data Streaming, Building Automation, Cloud Computing.

1 INTRODUCTION

"If you can't measure it, you can't improve it." – Peter Drucker

Building energy consumption is poorly understood, with many not achieving their expected energy performance (Fedoruk, et al., 2015; Mallory-Hill & Gorgolewski, 2018); discrepancies between simulated (design-phase models) and measured energy use can range as high as 70-80% (De Wilde, 2014; Menezes, et al., 2012). The use of IoT devices has significantly improved this precision, for example (Kim, et al., 2015; Zibin, et al., 2016). The value of building monitoring is well-established to maintain efficient building operation. Beyond simple monitoring, machine learning predictive analytics offer significant benefit to facility owners and operators to anticipate significant issues. For example, equipment performance can be identified tracked through online fault detection and diagnosis (FDD) (Lan & Chen, 2007; Li & O'Neill, 2018) and can inform preventative

[1] BSc candidate, Dept. Computer Science, Ryerson University, Toronto, Canada, tony.misic@ryerson.ca
[2] Post-doctoral fellow, Dept. Architectural Science, Ryerson University, Canada, saragilani@gmail.com
[3] Associate Professor, Dept. Architectural Science, Ryerson University, Canada email@xx.yy

maintenance or planned replacement, increasing system life (Beghi, et al., 2016). Energy efficiency can also be improved using strategies such as model predictive control and online commissioning; Smart and Continuous Commissioning (SCCx) has demonstrated significant potential for energy management, but commercial applications are limited (Verhelst, et al., 2017).

While also controlling equipment using prescriptive logic, Building Automation Systems (BAS) allow building operators to review the performance of individual systems while alerting them to alarms. Together, these control and monitoring functions work to ensure smooth functioning of the building. At present, most BAS software lacks the ability to perform the complex analysis necessary for FDD and SCCx, instead displaying only alarms and trends, and even "Smart" building systems available commercially are primarily limited to data mining and historical analysis. As such, a separate analytics platform is necessary to support these more complex applications, particularly for buildings with legacy systems. This paper responds to this need by presenting an approach for data ingestion and pre-processing of data streamed from a traditional BAS. The implementation details are presented as both a generalized approach as well as applied to a large mixed-use building on a university campus, which opened in Fall 2019. The planned analytics and expected stakeholder benefits are also discussed.

2 Background

A BAS consists sensors, actuators, and controllers on a dedicated network. Local equipment controllers (field controllers) constantly monitor sensor (point) values and, at either a prescribed change-of-value (COV) threshold or sampling frequency, sends updated point values through a branch of the BAS (trunk) to their respective network device (network automation engine) to the central BAS workstation. To facilitate the connection of a large number of devices, open protocols – most notably BACNet IP (ASHRAE, 2005) – are used to communicate between the BAS and third-party equipment. From the field controllers to the central workstation, proprietary protocols are typically used for communication along a dedicated network. A sample portion of a typical network architecture is shown in Figure 1.

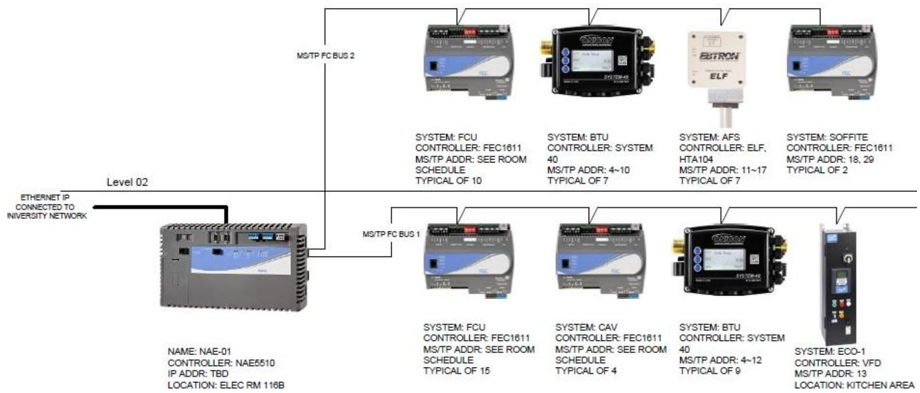


Figure 1 Example of a portion of a BAS architecture [12]

With the advent of IoT, there has been a significant interest in data streaming from sensor networks, for example (Ramprasad, et al., 2018), as well as the type of summary information relevant to the FM-BIM (Kassem, et al., 2015). The use of streamed BAS data for predictive analytics requires several key scientific challenges to be overcome: (1)

developing data definitions and structure; (2) development of a robust and secure approach for streaming the high-volume and heterogeneous BAS data; (3) development of metrics and key performance indicators; and (4) the training and implementation of analytics algorithms to monitor and calculate these metrics. The first two elements are within this paper's scope and key literature is summarized here.

Two approaches are used to structure data from sensor networks: linked data, and ontologies. Linked data stores individual the data from individual points separately, relying on an external software such as BIM to integrate this data into a common environment. Ontologies provide robust data structures and several broad ontologies have been developed to integrate BAS and IoT data (Bajaj, et al., 2017; Bhattacharya, et al., 2015), however these require that the BAS integrate significant semantic tags, which has yet to be implemented in the majority of systems.

Python scripts are widely used for computer networking applications due to its simplicity in coding and human interpretation, allowing complex or tedious network tasks to be automated quickly. Writing to an ElasticSearch (ES) cluster provides the end-user with flexibility in data interpretation. No extra interpretation must be written after processing, saving both time and resources. Furthermore, ES has multiple libraries in multiple languages, allowing for accessibility and scalability in future uses. One drawback is the indirect approach to using ES and the overhead computation when streaming into the cluster.

Cybersecurity is an increasing concern within the building sector, particularly for IoT applications as each new device provides a new potential point of entry for a data breach or cyberattack. BAS have traditionally not been designed to consider data security, relying instead on their presence on closed networks (Peacock, 2014). By breaching this network to permit streaming to an analytics server, vulnerabilities are introduced into this system that remain an open problem to be addressed, particularly given the high level of trust and limited data integrity checks inherent in these systems (Baig, et al., 2017).

3 IMPLEMENTATION

There are five components in the data streaming architecture: the BAS itself, a custom script developed in collaboration with the controls vendor to read the (proprietary) BAS network data and export it as text, a Python server to pre-process the data, the VPN connection, and the ES Cluster where future analytics will be performed.

3.1 Data Acquisition

There are two components in the data acquisition system: the BAS, and the software used to extract the BAS data and export it in a non-proprietary format. BAS systems vary by vendor, but typically follow the standard architecture described in Section 2.

It is critical to avoid overloading the BAS network with queries from the streaming system as this could result in lag on critical controls or – in the worst case – a BAS network failure. To avoid this, the BAS data acquisition must be a read-only system that 'sniffs' the data as it travels across the network and records it in the desired format. Unless this is done downstream of each field controller where BACNet IP is used for communication with individual equipment, this will be encoded in the proprietary BAS system, requiring coordination with the controls vendor to obtain a 'back door' to output these values in a pre-agreed, parsable text format. In this approach, COV outputs are embedded into the Data field (TCP body message) of one or more TCP packets (Fig. 2). The remainder of packet information is populated using standard TCP protocol rules (Dordal, 2014) and the completed packet is sent to the location coded into each packet to the Python server via the internal network. TCP does not permit confirmation to be sent unless all data is received in the correct order (Dordal, 2014), thus ensuring data integrity through the streaming process.

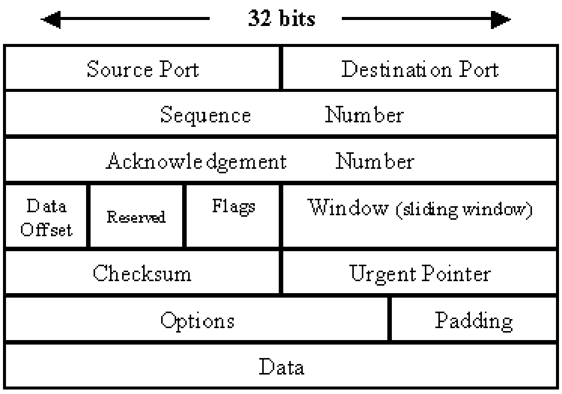


Fig. 2. TCP packet structure (Salomon, 2006)

3.2 Data Streaming

Data-streaming TCP packets is done through a running Python script on a physical machine located on the BAS ethernet network that points to an ES cluster index. This is shown schematically in Figure 3.

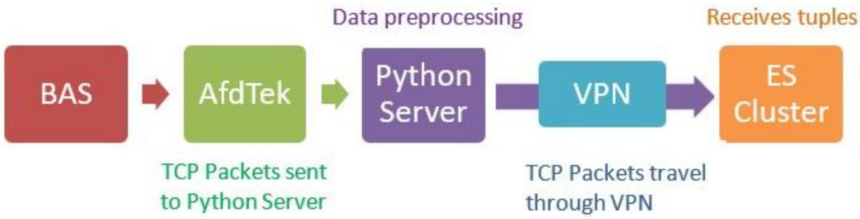


Fig. 3. Schematic of data streaming architecture from BAS to ES Cluster.

Python is used to create an indirect line of communication from the BAS to ES via the VPN, while also pre-processing the data. While the direct connection of the BAS head-end to the cloud is preferable from a reliability standpoint as it eliminates a potential point of failure – for example risk of disconnected cables, power failure at this intermediate computer, or system restarts – this permits a one-way connection from the BAS to the cloud, which is desirable from a facility management standpoint as it minimizes the risk of cyberattack. To ensure the system security on the BAS, we use a one-way output on the physical machine connected to the network. This only allows confirmation packets to come through the port securing the physical machine against attacks against its output port. This can lead to some lagged results but the information is eventually sent and therefore not a large issue. Further, this is configured as a one-way (read-only) connection at the cloud to further mitigate intrusion risk.

The script is assigned a predetermined IP address and port and that location information is inputted into the BACnet UI. The COV data sent by the BAS network devices to the Python server script, extracted from the TCP body message, and parsed using the .split and .strip functions in Python to populate a tuple consistent with the streamed data requirements discussed previously. Once the tuple has been populated with the appropriate values, the script attempts to write the data into the ES cluster index. Three means are used to avoid data loss. First, the IDs of the documents in the index are an Md5 hash (Salomon, 2006), a type of cryptographic key, is generated at each push attempt, minimizing the chance of duplicate ID's and avoiding non-duplicate data writing. Second, a backup file is written to each time a push attempt fails. In the process of writing from the file, indexing to ES is verified. This technique eliminates the loss of data during a downtime of the ES cluster or a client/server error. Finally, the entire server/client system is run over TCP/IP, which will guarantee information security and will not receive any incomplete data. Cybersecurity is enforced with a VPN, which connects the physical machine to the ES cluster.

3.3 Data Pre-Processing

The ES cluster will permit the structuring and pre-processing of data prior to storage in the data lake and facilitate queries by the future analytics. ES is extremely flexible, more so than SQL databases, and can be used it as a backend for a search engine, database, and analytics engine.

BAS point naming conventions, particularly in legacy systems, are frequently a) not human-readable, b) indicate the system and equipment measured but not the network location, or c) indicate the network location of the point but not the equipment or system being measured. To overcome this, a linked data approach is used to increase the context of each point the network location (NAE, trunk, and field controller), system and equipment identity, and point type information. This is achieved through a python script that calls a lookup table containing this data alongside the data point, adds it to each data instance tuple, and then parses the data to permit it's mapping to the data structure. By mapping the full context (unique) points, only a single search is necessary, and parsing is achieved by the structuring of the lookup tables.

Two nomenclatures are used to record the sensor network topology. A network context nomenclature in the form NAEDeviceID.TrunkID.FieldControllerID.PointType and a system context nomenclature that can be readily mapped to a FM-enabled BIM (as the third field is the host family instance) and asset management database in the form BuildingID.SystemID.EquipmentID.PointType.

4 CASE STUDY

The Daphne Cockwell Health Science complex (Fig 4).is a mixed-use building consisting of a 16,300m2 (175,000sf) academic podium that is primarily lab space and academic offices with 19-storey residence tower housing 332 student rooms in 2- and 4-bedroom apartments. Developed by the university to be a 'living lab', the building "contains a comprehensive sub-metering system that collects real-time data about energy consumption and climate control. The data can be used to identify opportunities to improve sustainability and inform decisions for future buildings, as well as being used for graduate-level research" (Ryerson University, 2019).

Fig. 4. Daphne Cockwell Complex at Ryerson University (Perkins + Will, 2019).

As noted, a 'back door' was implemented by the controls vendor (Johnson Controls Inc.) in collaboration with a subconsultant (AFDTek), permitting output of the timestamp, full-context point name, and value in text (TCP message) format from the BAS head-end. As noted in the literature review, the selection of TCP/IP for communication ensures data integrity, while the Md5 hash identifier prevented duplicate data.

System security is vital to the architecture avoiding cyberattacks on the host network and cloud and this was a key concern of the University. To address this, a secure VPN connection to a dedicated ES Cluster was used, configured to only permitting only the confirmation of TCP packet arrival to be sent back to the Python server (Fig. 5), to minimize the risk of cyber-intrusion. Figure 6 shows a sample of the same data as received by the ES Cluster.

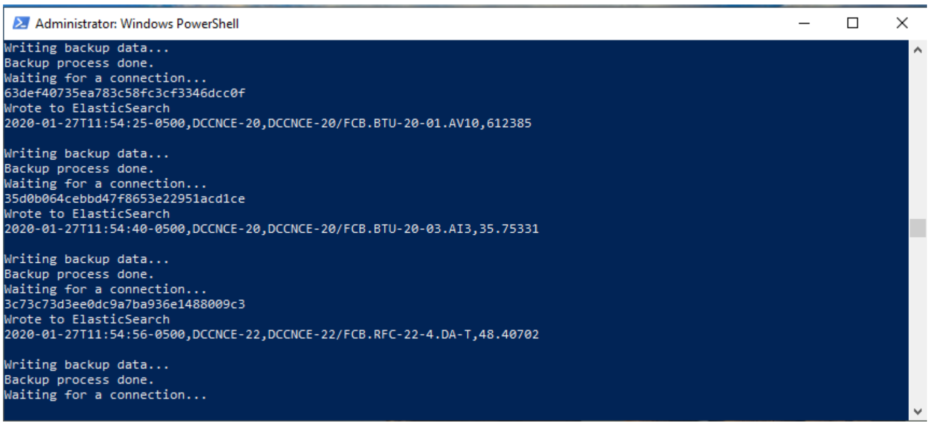
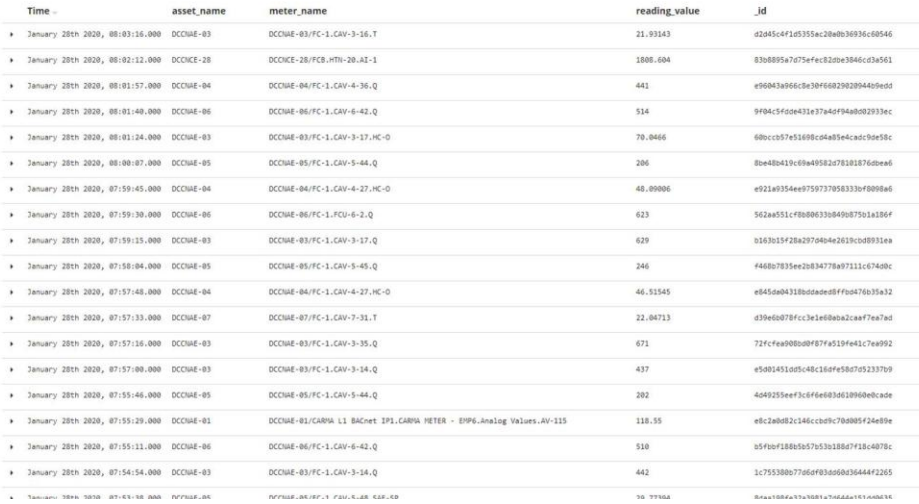


Fig. 5. Window showing data streaming to ES server via VPN



Time	asset_name	meter_name	reading_value	_id
January 28th 2020, 08:03:16.000	DCCNAE-03	DCCNAE-03/FC-1.CAV-3-16.T	21.93143	d2d45c4f1d5355ac20a0b36936c60546
January 28th 2020, 08:02:12.000	DCCNCE-28	DCCNCE-28/FCB.HTN-20.AI-1	1808.604	83b8895a7d75efec82dbe3846cd3a561
January 28th 2020, 08:01:57.000	DCCNAE-04	DCCNAE-04/FC-1.CAV-4-36.Q	441	e96043a966c8e30f66029020944b9edd
January 28th 2020, 08:01:40.000	DCCNAE-06	DCCNAE-06/FC-1.CAV-6-42.Q	514	9f04c5fdde431e37a4df94a0d02933ec
January 28th 2020, 08:01:24.000	DCCNAE-03	DCCNAE-03/FC-1.CAV-3-17.HC-O	70.0466	60bccb57e51698cd4a85e4cadc9de58c
January 28th 2020, 08:00:07.000	DCCNAE-05	DCCNAE-05/FC-1.CAV-5-44.Q	206	8be48b419c69a49582d78101876dbea6
January 28th 2020, 07:59:45.000	DCCNAE-04	DCCNAE-04/FC-1.CAV-4-27.HC-O	48.09006	e921a9354ee9759737058333bf8098a6
January 28th 2020, 07:59:30.000	DCCNAE-06	DCCNAE-06/FC-1.FCU-6-2.Q	623	562aa551cf8b80633b849b875b1a186f
January 28th 2020, 07:59:15.000	DCCNAE-03	DCCNAE-03/FC-1.CAV-3-17.Q	629	b163b15f28a297d4b4e2619cbd8931ea
January 28th 2020, 07:58:04.000	DCCNAE-05	DCCNAE-05/FC-1.CAV-5-45.Q	246	f468b7835ee2b834778a97111c674d0c
January 28th 2020, 07:57:48.000	DCCNAE-04	DCCNAE-04/FC-1.CAV-4-27.HC-O	46.51545	e845da04318bddaded8ffbd476b35a32
January 28th 2020, 07:57:33.000	DCCNAE-07	DCCNAE-07/FC-1.CAV-7-31.T	22.04713	d39e6b078fcc3e1e60aba2caaf7ea7ad
January 28th 2020, 07:57:16.000	DCCNAE-03	DCCNAE-03/FC-1.CAV-3-35.Q	671	72fcfea908bd0f87fa519fe41c7ea992
January 28th 2020, 07:57:00.000	DCCNAE-03	DCCNAE-03/FC-1.CAV-3-14.Q	437	e5d01451dd5c48c16dfe58d7d52337b9
January 28th 2020, 07:55:46.000	DCCNAE-05	DCCNAE-05/FC-1.CAV-5-44.Q	202	4d49255eef3c6f6e603d610960e0cade
January 28th 2020, 07:55:29.000	DCCNAE-01	DCCNAE-01/CARMA L1 BACnet IP1.CARMA METER - EMP6.Analog Values.AV-115	118.55	e8c2a0d82c146ccbd9c70d005f24e89e
January 28th 2020, 07:55:11.000	DCCNAE-06	DCCNAE-06/FC-1.CAV-6-42.Q	510	b5fbbf188b5b57b53b188d7f18c4078c
January 28th 2020, 07:54:54.000	DCCNAE-03	DCCNAE-03/FC-1.CAV-3-14.Q	442	1c755380b77d6df03dd60d36444f2265

Fig. 6. Raw storage of events as recorded on ES Cluster

In this case study, the streamed data output from the sniffer is in the format {Timestamp, NAEID, full context point name, value}. The full context point name in this instance is the network context point name described previously. Adding the system context point name was performed using a lookup table (Table 1). This system context point name is used internally on the BAS and uses the nomenclature BuildingID.SysID.BASID.PointID. For monitored equipment, the SysID is the building system containing the equipment, for example, the chilled water system, and BASID is the actual equipment being controlled, for example a chiller or chilled water pump. For room-based points, SysID is assigned "RM" and BASID is the room number, thus permitting a consistency in nomenclature that facilitates parsing into the SQL database. Taking in the formatted data a pre-processing Python script maps each column from the given data to the system context point IDs, which map each sensor, actuator, and controller to the appropriate equipment and system as well as network location.

Table 1. Lookup Table (selected rows)

Network Context PointID	System Context PointID
DCCNAE-01/FC-1.CAV-1-10.CLGUNOCC-SP	DCC.RM.DCC01-13.CLGUNOCC-SP
DCCNAE-01/FC-1.CAV-1-10.EFF-OCC	DCC.RM.DCC01-13.EFF-OCC
DCCNAE-01/FC-1.CAV-1-10.EFFCLG-SP	DCC.RM.DCC01-13.EFFCLG-SP
…	…

This system is being used to develop Continuous Commissioning and online optimization algorithms for deployment in the DCC building, in collaboration with the Facility Engineer. The scope of these planned algorithms includes online optimization and control of equipment and systems, for example predicting the ideal chilled and condenser water temperature setpoints to minimize chiller system energy consumption, and fault detection and diagnosis, for example tracking chiller fouling.

5 Discussion and Conclusions

This paper has presented a standard approach to data acquisition from a building automation system, which could be applied to any IoT sensor network, and the necessary processing required to structure and stream this data. This approach has been applied in a large academic building with over 10,000 BAS points and has demonstrated itself to be robust in implementation.

This approach has significant value to support Smart Continuous Commissioning and other online controls optimization efforts. While advertised by many vendors, existing BAS “smart” analytics capabilities are limited to data visualization and the few vendor solutions available require additional investment beyond the cost of the BAS Moreover, this functionality is typically a "black box" application where the Facility Engineer can neither inspect the algorithm to understand what is being considered and how the controls are being changed as a result. As was the case in this study, such an approach can result from a lack of trust in these applications. An open analytics system, on the other hand, addresses these issues of transparency while also providing the Facility Engineer with flexibility in modifying these to better suit their unique needs.

The streaming of BAS data facilitates several other valuable stakeholder activities. For example, this data can be readily mapped into an FM enabled BIM, where it can be overlaid with data integrated from other FM systems such as preventative maintenance, space management, call center logs and document complaints, building inspection records, et cetera. The nomenclature presented, whether built into the BAS naming conventions for new systems or added to the streamed data during pre-processing, further supports FM-enabled BIM by embedding significant semantic information regarding system and sensor network architecture to each point. For buildings where no BIM readily exists, a lightweight FM-enabled BIM can be rapidly developed (for example, using the approaches presented in (McArthur & Bortoluzzi, 2018; Bortoluzzi, et al., 2019)), and this information can be mapped to the equipment to quickly add the building systems topology.

The presence of building identification tags for the data supports multi-building contexts, ensuring unique point identifiers across campuses and building portfolios, and thus enabling the nesting or linking of multiple FM-enabled BIMs to generate full-site visualizations.

The primary limitation of this paper is that only a single implementation has been tested. Use a specific system, the JCI medicine system and made use of a vendor provided back door to access the text data. An equivalent means of of sniffing the change of value data from another BAS is necessary for the supplementation.

Future work will include the development of key performance indicators for the facility and their optimization. For example, the first priority KPI to be developed is chiller plants energy consumption, and the next stage and the first set of analytics to be developed will monitor the chiller system infer its performance and required characteristics and identify the optimal controls points to minimize system use chiller energy use.

6 Acknowledgments

This research was funded by the Natural Sciences and Engineering Research Council of Canada [CRDPJ 461929 - 13], the Ontario Research Fund [RE07-023], and FuseForward Solutions Group.

7 References

Fedoruk, L. Cole, R. Robinson, J., Cayuela, A.: Learning from failure: understanding the anticipated–achieved building energy performance gap. Building Research and Information 43(6), pp. 750-763 (2015).

Mallory-Hill, S., Gorgolewski, M.: Mind the Gap: Studying Actual Versus Predicted Performance of Green Buildings in Canada. Building Performance Evaluation, pp. 261-274 (2018).

De Wilde, P.: The gap between predicted and measured energy performance of buildings: a framework for investigation. Automation in Construction 41, pp. 40-49 (2014).

Menezes, C., Cripps, A., Bouchlaghem, D., Buswell, R.: Predicted vs. actual energy performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap. Applied Energy 97, pp. 335-364 (2012).

Kim, J.B., Jeong, W., Clayton, M.J., Haberl J.S., Yan, W.: Developing a physical BIM library for building thermal energy simulation. Automation in Construction 50, pp. 16-28 (2015).

Zibin, N., Zmeureanu, R. Love, J.: Automatic assisted calibration tool for coupling building automation system trend data with commissioning. Automation in Construction 61, pp. 124-133 (2016).

Lan, L., Chen, Y.: Application of Modeling and Simulation in Fault Detection and Diagnosis of HVAC Systems. In: Proceedings of Building Simulation 2007, pp. 1299-1306, IBPSA, Beijing (2007).

Li, Y., O'Neill, Z.: A critical review of fault modeling of HVAC systems in buildings. Building Simulation *VOL", pp. 953-975 (2018).

ASHRAE: Standard 135-2004, BACnet-A Data Communication Protocol for Building Automation'and Control Networks, American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta (2005).

Johnson Controls Inc.: DCC BAS Shop Drawings. Toronto (2018).

Ramprasad, B., McArthur, J., Fokaefs, M., Barna, C., Damm, M., Litoiu, M.: Leveraging existing sensor networks as IoT devices for smart buildings. In: Proceedings of the Global Forum on Internet of Things, IEEE, Singapore (2018).

Kassem, M., Kelly, G., Dawood, N., Serginson, M., Lockley, S.: BIM in facilities management applications: a case study of a large university complex. Built Environment Project and Asset Management 5(3), pp. 261-277 (2015).

Bajaj, G., Agarwal, R., Singh, P., Georgantas, N., Issarny, V.: A study of existing Ontologies in the IoT-domain. arXiv preprint arXiv:1707.00112 (2017).

Bhattacharya, A., Ploennigs, J., Culler, D.: Short Paper: Analyzing metadata schemas for buildings: The good, the bad, and the ugly. In: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, Seoul, pp. 33-34, (2015).

Peacock, M.J.M.: An analysis of security issues in building automation systems. In: Proceedings of the 12th Australian Information Security Management Conference, Perth (2014).

Baig, Z., Szewczyk, P., Valli, C., Rabadia, P., Hannay, P., Chernyshev, M., Johnstone, M., Kerai, P., Ibrahim, A., Sansurooah, K., Syed, N.: Future challenges for smart cities: Cyber-security and digital forensic. Digital Investigation 22, pp. 3-13 (2017).

Dordal, P. L.: An Introduction to Computer Networks, McGraw-Hill Primis Custom Publishing (2014).

Salomon, D.: Foundations of Computer Security, Northridge (2006).

Introducing the Daphne Cockwell Health Sciences Complex at Ryerson, https://www.ryerson.ca/news-events/news/2019/11/introducing-the-daphne-cockwell-health-sciences-complex-at-ryerson/, last accessed 20 01 2020.

Ryerson University Daphne Cockwell Health Sciences Complex, https://perkinswill.com/project/daphne-cockwell-health-sciences-complex/, last accessed 21 01 2020.

McArthur, J., Bortoluzzi, B.: "Lean-Agile FM-BIM: a demonstrated approach. Facilities 36(13/14), pp. 676-695 (2018).

Bortoluzzi, B., Efremov, I., Medina, C., Sobieraj, S., McArthur, J.: Automating the creation of building information models for existing buildings. Automation in Construction 105, p. 102838 (2019).

Schluse, M., Priggemeyer, M., Atorf, L., Rossmann, J.J.: Experimentable digital twins—Streamlining simulation-based systems engineering for industry 4.0. IEEE Transactions on Industrial Informatics 14(4), pp. 1722-1731 (2018).

Kučera, A., Pitner, T.: Semantic BMS: Ontology for Analysis of Building Automation Systems Data," In: 7th Doctoral Conference on Computing, Electrical and Industrial Systems (DOCEIS), Costa de Camperica, publisher (2016).

Building Performance Institute Europe. Building Automation and Control Technologies (Report No. 4). 2016. [Online]. Available: http:/bpie.eu.

Janowicz, K., Haller, A., Cox, S., Le Phuoc, D., Lefrançois, M.: SOSA: A lightweight ontology for sensors, observations, samples, and actuators," Journal of Web Semantics 56, pp. 1-10 (2019).